

Artificial neural networks forecasting of PM_{2.5} pollution using air mass trajectory based geographic model and wavelet transformation



Xiao Feng*, Qi Li, Yajie Zhu, Junxiong Hou, Lingyan Jin, Jingjie Wang

Institute of Remote Sensing and GIS, Peking University, Beijing 100871, China

HIGHLIGHTS

- We propose a novel hybrid model to forecast PM_{2.5} pollution.
- Using trajectory based geographic parameter as an extra input to ANN model.
- Applying prediction strategy at different scales and then sum them up.
- The model is capable to predict the high peaks of PM_{2.5} concentrations.

ARTICLE INFO

Article history:

Received 26 November 2014

Received in revised form

7 February 2015

Accepted 10 February 2015

Available online 11 February 2015

Keywords:

PM_{2.5} forecasting

Artificial neural networks

Air mass trajectory based geographic model

Wavelet transformation

ABSTRACT

In the paper a novel hybrid model combining air mass trajectory analysis and wavelet transformation to improve the artificial neural network (ANN) forecast accuracy of daily average concentrations of PM_{2.5} two days in advance is presented. The model was developed from 13 different air pollution monitoring stations in Beijing, Tianjin, and Hebei province (Jing-Jin-Ji area). The air mass trajectory was used to recognize distinct corridors for transport of “dirty” air and “clean” air to selected stations. With each corridor, a triangular station net was constructed based on air mass trajectories and the distances between neighboring sites. Wind speed and direction were also considered as parameters in calculating this trajectory based air pollution indicator value. Moreover, the original time series of PM_{2.5} concentration was decomposed by wavelet transformation into a few sub-series with lower variability. The prediction strategy applied to each of them and then summed up the individual prediction results. Daily meteorological forecast variables as well as the respective pollutant predictors were used as input to a multi-layer perceptron (MLP) type of back-propagation neural network. The experimental verification of the proposed model was conducted over a period of more than one year (between September 2013 and October 2014). It is found that the trajectory based geographic model and wavelet transformation can be effective tools to improve the PM_{2.5} forecasting accuracy. The root mean squared error (RMSE) of the hybrid model can be reduced, on the average, by up to 40 percent. Particularly, the high PM_{2.5} days are almost anticipated by using wavelet decomposition and the detection rate (DR) for a given alert threshold of hybrid model can reach 90% on average. This approach shows the potential to be applied in other countries’ air quality forecasting systems.

© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Most major cities in the northeast of China, especially for Beijing, Tianjin and Hebei province, known as a rapidly developed agglomeration, have experienced severe short-term pollution events that are harmful to human health (e.g., Liu et al., 2013; Wang

et al., 2013). The dominant pollutant, particularly PM_{2.5} (particulate matter with aerodynamic diameter below 2.5 μm) in haze pollution, is epidemiologically associated with the risk of deleterious health effects on cardiovascular and lung diseases (e.g., Du et al., 2010; Qiu et al., 2013). Well-documented adverse effects of PM_{2.5} have stimulated intensive research targeting at simulating and forecasting its behavior. However, the wide spread sources of PM_{2.5}, including industrial process, energy production from power stations, vehicular traffic, residential heating, transport, natural disasters, coupled with the complicated physical and chemical

* Corresponding author.

E-mail address: fengxiao198995@163.com (X. Feng).

processes, make the PM_{2.5} forecasting a difficult task. Several types of approaches have been used, fundamentally branching into two main streams: deterministic and statistical approaches.

Deterministic approaches can be performed without a large quantity of historical data, but it demands sufficient knowledge of pollutant sources, the real-time emission quantity, the explicit description of major chemical reactions among exhaust gasses and temporal physical processes under the planetary boundary layer (PBL). Recent deterministic forecasts (e.g., McKeen et al., 2009; Chuang et al., 2011) are elaborated by online/offline-coupled meteorological-chemistry models that are composed of simplified or more comprehensive 3-D chemistry transport models (CTMs). The use of coupled 3-D CTMs would significantly enhance the routine forecasts, thus promoting the understanding of the underlying complex interaction between meteorology, chemistry, and emission. Although CTMs based forecasts are capable to predict spatially resolved concentrations in places without monitoring site, the key knowledge is often insufficient and in some cases, it is computationally expensive. Approximations and simplifications are, therefore, often employed in the processing of CTMs. Limited knowledge of pollutant sources and imperfect representation of physicochemical processes would pose rather strong biases in forecasted concentrations (Stern et al., 2008).

On the other hand, statistical approaches often require a large quantity of historical measurement data under various atmospheric conditions. Different functions can be used to establish the respective relationships between the routinely-measured pollutant data and the various selected predictors using regression and machine learning methods. The major drawback of this approach is that it is the best representative of only a specific monitoring station and cannot be extended to other regions with different meteorological conditions (Niska et al., 2005). Nevertheless, the statistical approach is generally more appropriate for the discovering of underlying complex site-specific dependencies between concentrations of air pollutants and potential predictors (Hrust et al., 2009), and consequently, they often have a higher accuracy, as compared with deterministic models. The commonly-used statistical approaches include multiple linear regression (MLR) (e.g., Stadlober et al., 2008; Genc et al., 2010), ANNs (e.g., Pérez and Reyes, 2006; Li and Hassan, 2010), support vector machine (SVM) (e.g., Guyon et al., 2002; Osowski and Garanty, 2007), fuzzy logic (FL) (e.g., Shad et al., 2009; Alhanafy et al., 2010), Kalman filter (KF) (e.g., Zolghadri and Cazaurang, 2006; Hoi et al., 2008) and hidden Markov model (HMM) (Sun et al., 2013). Some studies (e.g., Gardner and Dorling, 1999) have suggested that the interplay of human, climate, and air pollution is too complex to be represented in deterministic models without developing a separate statistical model. Evidence has shown that ANNs can simulate nonlinearities and interactive relationships, getting more accurate results than a CTM such as CHIMERE (e.g., Dutot et al., 2007). In spite of this, ANNs are supposed to be combined with other models in order to overcome their limitations (Díaz-Robles et al., 2008).

Artificial neural networks have been frequently used as a non-linear tool in recent atmospheric and air quality forecasting studies. Gardner and Dorling (1998) gave an informative review of the applications of ANN in atmospheric sciences. They pointed out the advantages of ANN when handling with non-linear systems, especially when theoretical models are difficult to be constructed. Perez and Reyes (2002) applied an MLP and linear model to predict the maximum of 24-h average PM₁₀ concentrations in Santiago, Chile. They argued that although MLP gives a slightly better result than linear model, the selection of potential predictors is more important than the model types (MLP or linear). Kukkonen et al. (2003) evaluated five ANN models in comparison to one linear and one deterministic model for the prediction of PM₁₀ and NO₂

concentrations in Helsinki, Finland. The ANN models gave better results than other models, especially for the ANN models that were built with non-constant variance. A special MLP model for the forecasting of daily average air quality index (API) was presented by Jiang et al. (2004). They modified the training method as well as the model structure of ANN and significantly improved the accuracy of prediction. The authors also suggested that a simpler structure of MLP models with early stop training gives better results on test data. Hooyberghs et al. (2005) improved the accuracy of forecasting daily averages of particulate matter by treating boundary layer height as one of the input variables in ANN model. They concluded that the meteorological conditions played a significant role in the day-to-day fluctuations of PM₁₀ concentrations in Belgium. Lu et al. (2006) employed a two-stage ANN to forecast ozone concentrations. The meteorological conditions were firstly clustered into different meteorological regimes using self-organizing map (SOM). Then, a supervised MLP model was used to approximate the nonlinear ozone-meteorological relationship in each meteorological regime. They found the hybrid SOM-MLP model can explain at least 60% of the variance in the ozone concentrations. Brunelli et al. (2007) investigated the applicability of recurrent neural network (Elman model) for the prediction of daily maximum concentrations of different pollutants. They found somewhat better consistency between forecasted and measured concentrations for Elman networks, as compared with MLP. Díaz-Robles et al. (2008) proposed a novel hybrid ARIMA-ANN model to improve the PM₁₀ forecast accuracy in Temuco, Chile. Experimental results showed that the hybrid model can effectively improve the forecasting accuracy obtained by either of the models used separately. An interesting approach of selecting the average intervals for input variables was attempted by Hrust et al. (2009). They selected optimal averaging periods for each potential predictor by comparing the values of correlation coefficient between modeled and measured concentrations. Sensitivity analysis for each input variable was also conducted in this experiment. Kurt and Oktay (2010) proposed a geographic based model to forecast the daily average concentrations of SO₂, CO and PM₁₀ three days in advance using MLP. They employed three kinds of geographic models: the single-site neighborhood model, the two-site neighborhood model and distance-based model. Experimental results showed that geographic based models outperformed the plain model, especially for the distance-based model. It is expected that there still have much space for improvement if more meteorological variables are added to the geographic model. Siwek and Osowski (2012) applied wavelet transformation with ANN ensemble to predict the daily average concentrations of PM₁₀. They combined several types of ANN in one ensemble to make a final prediction in an additional neural network. Results showed the usefulness of wavelet transformation in air pollution forecasts. A review of real-time air quality forecasting methods was given by Zhang et al. (2012a, 2012b).

A common drawback with these models, however, is that during the very high PM days, the forecasting errors tend to be much larger, and the PM concentrations are systematically underpredicted. It is these very events that pose the most adverse effects on our health. In order to capture the abrupt changes in PM concentrations with statistical approaches, some priori knowledge is needed. For instance, the characteristics of air pollution (local accumulation or transport) in selected areas. Feng et al. (2014) made a comprehensive analysis on the formation and dominant factors of haze pollution in Beijing. Air mass trajectory analysis was used to identify the corridors for transport of "dirty" air and "clean" air to Beijing. The aim of this research is to develop a hybrid model combining air mass trajectory analysis and wavelet transformation to improve the ANN forecast accuracy of daily average concentrations of PM_{2.5} two days in advance. The effectiveness of wavelet

transformation in time series analysis has been showed by some recent papers (Sharuddin et al., 2008; Zaharim et al., 2009; Siwek and Osowski, 2012).

2. Data and methods

2.1. Study area and available data

Beijing, Tianjin and Hebei province (Jing-Jin-Ji area), located in the northeast of China (Fig. 1a), have experienced a rapid increasing of urbanization. Urban population, energy consumption, increasing number of vehicles, has contributed to the situation of gradually exacerbation of atmospheric pollution (Marshall et al., 2008; Zhang et al., 2012). Since once occurred in notoriously foggy London, frequent episodes of regional haze pollution have occurred in major cities over northern China, especially in megalopolis agglomeration centralized with Beijing (Liu et al., 2013). Dominant northwestern winds in spring transported natural dust from Gobi deserts to the east of China, resulting in dust storms with the PM₁₀ concentration more than 1 mg/m³ (Liu et al., 2008). Besides, Jing-Jin-Ji area is surrounded by mountains on two sides, in the north the Jundu Mountain, part of Yanshan Mountains, and in the west the Xishan Mountain, part of Taihang Mountains, which makes the air pollutants difficult to be driven away.

As the Environmental Protection Administration (EPA) of China published new air quality standard in 2012 (EPA, 2012), networks of air pollution monitoring stations in major cities such as Beijing and Shanghai have been established. There are 80 air pollution monitoring sites in the area of Jing-Jin-Ji (marked by red points in Fig. 1b). The hourly concentrations data of air pollutants (PM_{2.5}, PM₁₀, NO₂, SO₂, O₃, CO) are obtained by each monitoring station automatically. At the beginning of 2013, the EPA of China gradually published these real-time data to the public. We collected a dataset for more than one year, ranging from 1 September 2013 to 31

Table 1

Statistics of measured values. Unit, range, maximum, minimum, mean, and standard deviation values from 1 September 2013 to 31 October 2014.

Variable	Unit	Range	Mean	St. Dev.
PM _{2.5}	µg/m ³	[6.25, 448]	90.27	74.33
Tmax (hourly)	°C	[-3, 42]	19.78	9.92
Tmin (hourly)	°C	[-11, 28]	9.65	9.69
Humidity	%	[12.91, 95.62]	54.28	19.76
Wind _x	series	[-2.86, 1.80]	0.15	0.73
Wind _y	series	[-2.12, 1.69]	-0.07	0.72
Day of year (DOY)	float	[-1, 1]	—	—
Day of week	int	[1, 7]	—	—
General condition	int	[1–10]	—	—

October 2014. The hourly meteorological forecast data including temperature, relative humidity, wind speed (varying from 1 to 12), wind direction, and general condition in county level (marked by green points in Fig. 1b) were obtained for the same period from the China Weather Website Platform, which is maintained by China Meteorological Bureau. Air pollution and meteorological data were matched by searching the nearest county/station from them. Due to the instrument malfunctions, some data were missing. For the purpose of model development, only records with both pollution and meteorological data were taken in to account. Days with consecutive hourly gaps of more than 4 h or the cumulative number of missing data exceeded 8 h were discarded. The final version for experiment is a dataset covering 404 days.

A statistical summary of measured variables from 1 September 2013 to 31 October 2014 is given in Table 1. The values (maximum, minimum, mean and standard deviation) are based on the average daily values in each monitoring station of Jing-Jin-Ji area. For the purpose of model development, wind direction was transformed into two components, southern and eastern. Thus, wind can be represented by two vectors with Eq. (1):

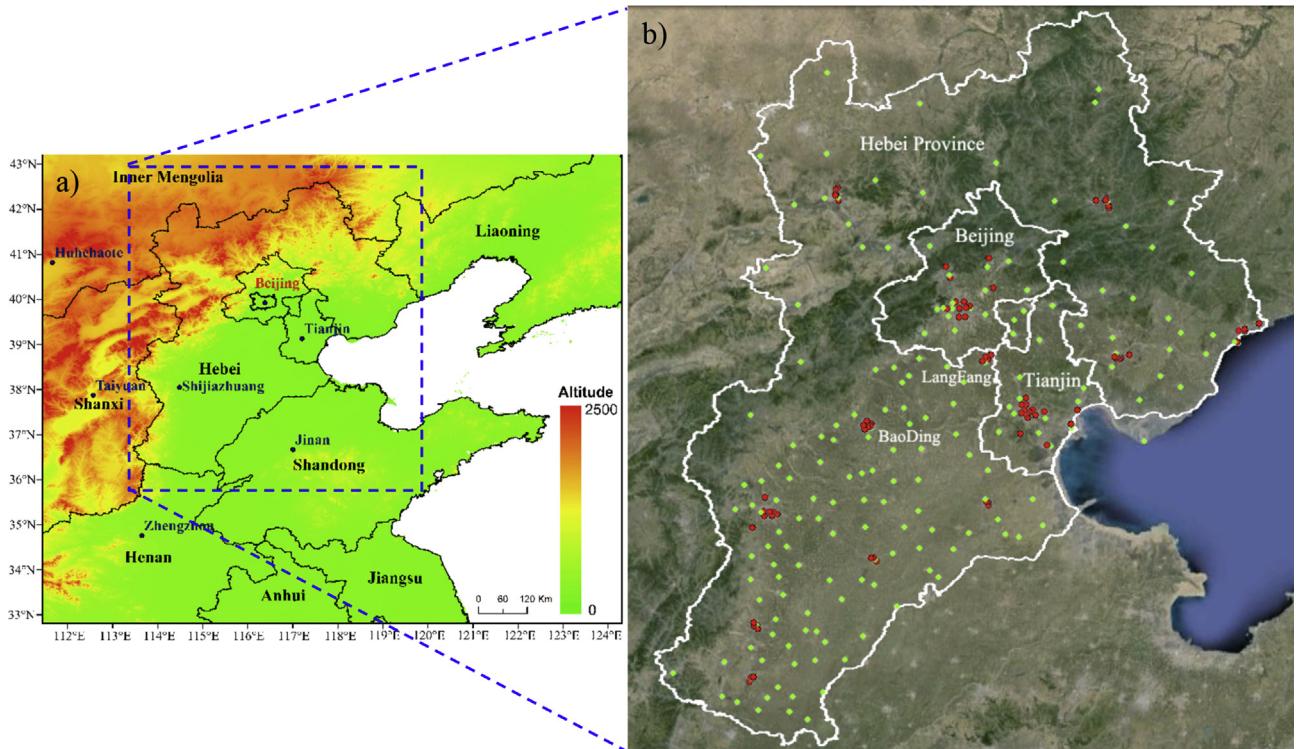


Fig. 1. a) The topography of surrounding areas of Beijing, Tianjin, and Hebei province (Jing-Jin-Ji area); b) Locations of air pollution monitoring stations (marked by red points) and county level meteorological forecasts (marked by green points). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$\begin{aligned} Wind_x &= w \cos \phi \\ Wind_y &= -w \sin \phi \end{aligned} \quad (1)$$

where w denotes the wind speed and ϕ is the angle of wind direction. The item day of week in the table varying from 1 to 7 with 1 corresponds to Monday. The day of year (DOY) is calculated by Eq. (2):

$$DOY = \cos(2\pi d_{th}/T) \quad (2)$$

where d_{th} represents the ordinal number of the day in the year and T is the number of days in this year. The value of DOY is confined to $[-1, 1]$ with a maximum in winter and a minimum in summer. Such representation guarantees the continuity of the input variables. The general condition (varying from 1 to 10) has 10 different weather conditions such as sunny, cloudy, rainy, etc. Daily average concentrations of PM_{2.5}, humidity, and daily maximum and minimum temperatures are also presented in Table 1.

To develop a good prediction model, we must generate the proper input prognostic features. So the relationship between PM_{2.5} and other meteorological variables is quite important. It is suggested to avoid the meteorological parameters strongly correlated with each other (Kuncheva, 2004). Table 2 shows the values of correlation coefficients. It is found that all the meteorological variables are weakly correlated with each other except for daily maximum and minimum temperatures. These two variables are quite crucial in the daily average prediction, for they represent the coldest and warmest condition of a day. Thus, there is no reason to drop any of them.

2.2. Air mass trajectory based geographic model

Air mass trajectory analysis has been a useful tool for detecting the direction and location of sources for various air pollutants. For the purpose of improving the PM_{2.5} forecast model, we tried to identify the transport patterns, which were associated with high PM_{2.5}. Backward trajectories from Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model were used to track the transport corridors of air masses that arriving at Beijing (Feng et al., 2014). The HYSPLIT model computes the trajectories with gridded wind field data, which are based on interpolations of rawwindsonde data (Draxler, 1991). Using trajectory clustering technique, we have tracked three transport corridors in Jing-Jin-Ji area (Fig. 2): one extending southwest, and approximately centered alone the industrial areas in the southwest of Hebei province (Fig. 2a), one extending southeast to Tianjin Municipality (Fig. 2b), and one extending to the northwest, reaching the Gobi deserts of Inner Mongolia (Fig. 2c). These corridors considered to be a “region of influence” for transport of “dirty” or “clean” air into Beijing. The spatial distribution of yearly average PM_{2.5} concentrations in Jing-Jin-Ji area was obtained by applying the interpolation method of Kriging (Fig. 3). The three corridors in Fig. 2 correspond to the three distinct regions marked with circles. The two high PM regions in the southwest and southeast of Beijing

were marked with black circle indicating the “dirty” air source, while the low PM region in the northwest of Beijing was marked with green circle indicating the “clean” air source.

The trajectory based geographic model was developed on the idea of using three or more neighboring sites’ air pollution indicator value as an additional input for the current station. The selection of neighboring stations should be primarily based on the distribution of transport corridors. Another important factor to consider is the distances between neighboring sites. In this study, we selected four air pollution monitoring sites (A, B, C, D in Fig. 3) to test the effectiveness of this model. The descriptions of the four sites are presented in Table 3. Station A and B locate in the vicinity of “dirty” air corridors while station C locates by the “clean” air corridor. Station D is in the central urban area of Beijing, which is surrounded by station A, B, and C. For each monitoring site, a triangular station net was constructed based on the transport corridors and the distances between the neighboring sites. Distance values were used to calculate a weighted average of air pollutant concentrations in three neighboring stations. Particularly, when the wind speed is greater than zero, wind direction should be considered as another factor to revise the weights calculated by distances. According to the wind direction at the station being predicted, we adjust the weights at three selected neighboring stations by adding weights to the upwind stations and reducing the weights of downwind stations. Stations adjust their weights according to the proportion of their inverse distance weight to ensure a zero delta in total. For example, if station C is the upwind station, we can use the Eq. (3) to compute the extra input of station D.

$$\begin{aligned} \bar{C}_D = C_A * \left(W_{AD} - \frac{\delta * I * W_{AD}}{W_{AD} + W_{BD}} \right) + C_B * \left(W_{BD} - \frac{\delta * I * W_{BD}}{W_{AD} + W_{BD}} \right) \\ + C_c * (W_{CD} + \delta * I) \end{aligned} \quad (3)$$

In this expression, C_A represents the PM_{2.5} concentration at station A, W_{AD} is the normalized inverse distance weight (IDW) between station A and D ($W_{AD} + W_{BD} + W_{CD} = 1$), I denotes the wind speed, δ is a constant and it takes the value of 0.1 in this study. This constant is designed to adjust the weights calculated by distances at neighboring sites.

Thus, the trajectory based geographic model is considered capable to capture both atmospheric and geospatial information. It is expected that using the calculation result of the above equation as an extra input to neural networks cannot only increase the overall forecasting accuracy, but also to some extent solve the problem of the underprediction of high PM days.

2.3. Wavelet transformation of the time series

The aim of this task is to forecast the daily average concentrations of PM_{2.5} two days in advance using the previous day’s concentration and the predicted values of other meteorological variables for the day underprediction. Due to the severe air pollution in Jing-Jin-Ji area, the high variability of PM_{2.5} time series makes the accurate prediction a very difficult task. Another way to solve the problem is to decompose the original high variability time series into a few sub-series with lower variability, applying the designed neural network to each of them and then sum up the individual results. So in this study, we will use the wavelet transformation to conduct this work. Some other papers have shown the usefulness of wavelet transformation in time series analysis and prediction (Sharuddin et al., 2008; Zaharim et al., 2009; Siwek and Osowski, 2012).

Discrete wavelet transformation can decompose the time series

Table 2

The cross correlation coefficients between different air pollutant predictors.

	PM _{2.5}	Tmax	Tmin	Humidity	Wind _x	Wind _y
PM _{2.5}	1	-0.214	-0.130	0.231	0.101	-0.098
Tmax	-0.214	1	0.927	0.291	0.089	0.368
Tmin	-0.130	0.927	1	0.418	0.164	0.327
Humidity	0.231	0.291	0.418	1	0.529	0.264
Wind _x	0.101	0.089	0.164	0.529	1	0.295
Wind _y	-0.098	0.368	0.327	0.264	0.295	1

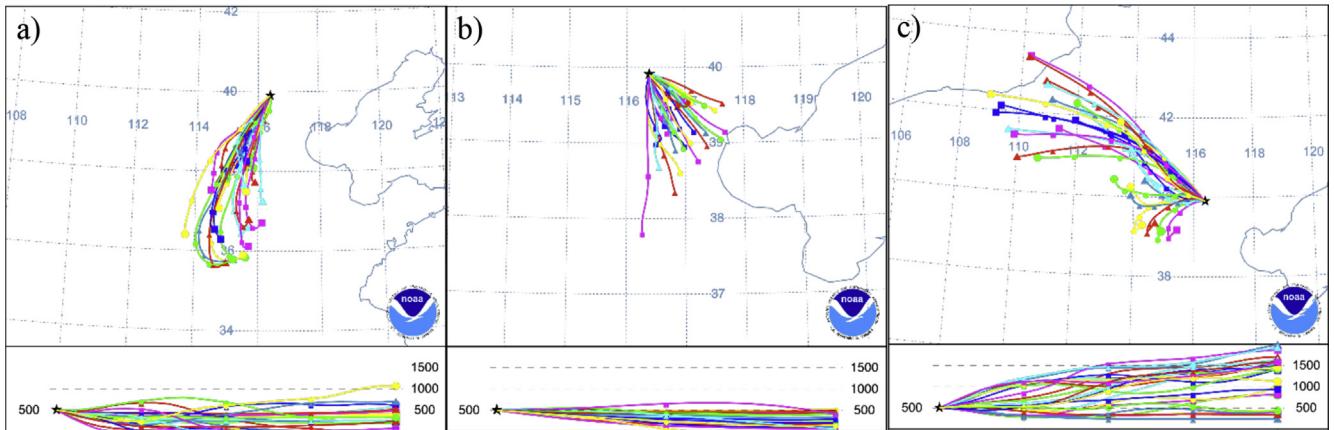


Fig. 2. Three transport corridors, namely, southwest branch (a), southeast branch (b), and northwest branch (c), tracked by 24 h backward trajectories of air masses in Jing-Jin-Ji area.

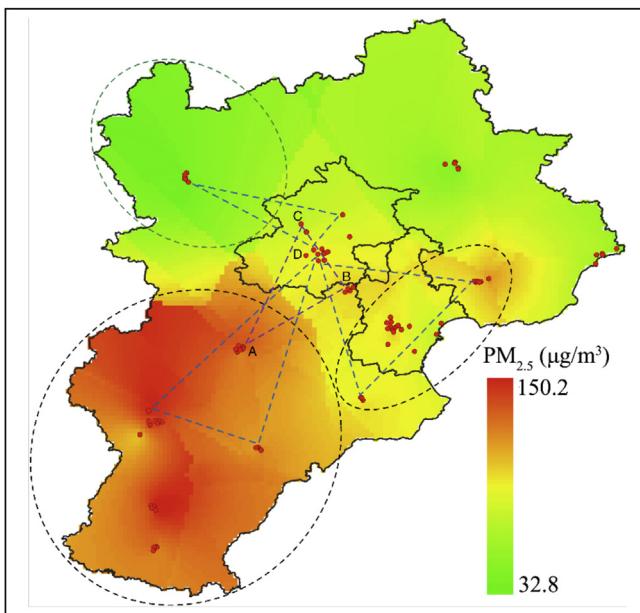


Fig. 3. The spatial distribution of yearly average PM_{2.5} concentrations in Jing-Jin-Ji area. The two high PM regions in the southwest and southeast of Beijing are marked with black circles, while the low PM region in the northwest of Beijing is marked with green circle. Four air pollution monitoring sites selected in the experiments are marked with A, B, C, and D. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

into a finite summation of shifted wavelets in different scales using Eq. (4) (Daubechies, 1988; Mallat, 1989):

$$s(n) = \sum_j \sum_k c_{jk} \psi(a_0^j nT - k) \quad (4)$$

Table 3
Descriptions of the four monitoring stations.

Site	Station name	Longitude	Latitude	Elevation (m)	Zone
A	HuaDianErQu	115.51 E	38.89 N	18.9	"dirty" source
B	HuanJingJianCeZhongXin	116.70 E	39.55 N	14.3	"dirty" source
C	DingLing	116.17 E	40.29 N	79.7	"clean" source
D	GuCheng	116.23 E	39.93 N	63.8	central urban

Where $s(n)$ represents the original time series, c_{jk} is a set of wavelet coefficients, $\psi(a_0^j nT - k)$ denotes the wavelet on j th scale shifted by k samples, a_0 is a constant, and normally it takes the value of 2 (Dyadic wavelet). In our approach, the original time series of PM_{2.5} was decomposed by discrete wavelet transformation into detailed wavelet coefficients $D_j(k)$ of the proper time shifts k in different scales ($j = 1, 2, \dots, 5$) and the coarse approximation signal $A_5(k)$.

Fig. 4 shows the results of the 5-level wavelet decomposition of the original time series of PM_{2.5} concentrations at station D. They were obtained by applying Db5 wavelets implemented in the wavelet toolbox of Matlab 2014a. Db5 was chosen as the wavelet function as it provided the smallest variability of time series at the particular levels. The optimal value of J was determined by the standard deviation of the approximation A_j . Osowski and Garanty (2007) gave an empirical expression on it shown by Eq. (5). For the data presented in Fig. 4, the ratio $std(A_5)/std(s) = 0.093$ satisfies the equation (5).

$$\frac{std(A_j)}{std(s)} < 0.1 \quad (5)$$

The final prediction of PM_{2.5} concentration at time t was then obtained by applying the Eq. (6):

$$PM_{2.5}(t) = f(D_1(t-1), \dots) + f(D_2(t-1), \dots) + \dots + f(D_5(t-1), \dots) + f(A_5(t-1), \dots) \quad (6)$$

where $D_j(t-1)$ represents the value of the individual time series of level j at time $t-1$, and function f denotes the mappings between the predicted concentration and various potential predictors.

2.4. Neural network architecture

In this part, we present the architecture of the ANN model used in predicting the daily average concentrations of PM_{2.5} two days in advance in the experiments. The architecture of this forecasting model is shown in Fig. 5. For this model, a multi-layer perceptron type of back-propagation neural network was chosen. Logistic sigmoid function was used as the transfer function in hidden layer. The Levenberg–Marquardt (LM) (Shepherd, 1997) algorithm was employed for training and the method of early stopping (Sarle, 1995) was used to avoid overfitting. Since the forecasting was conducted for two days in advance, the predicted concentration for

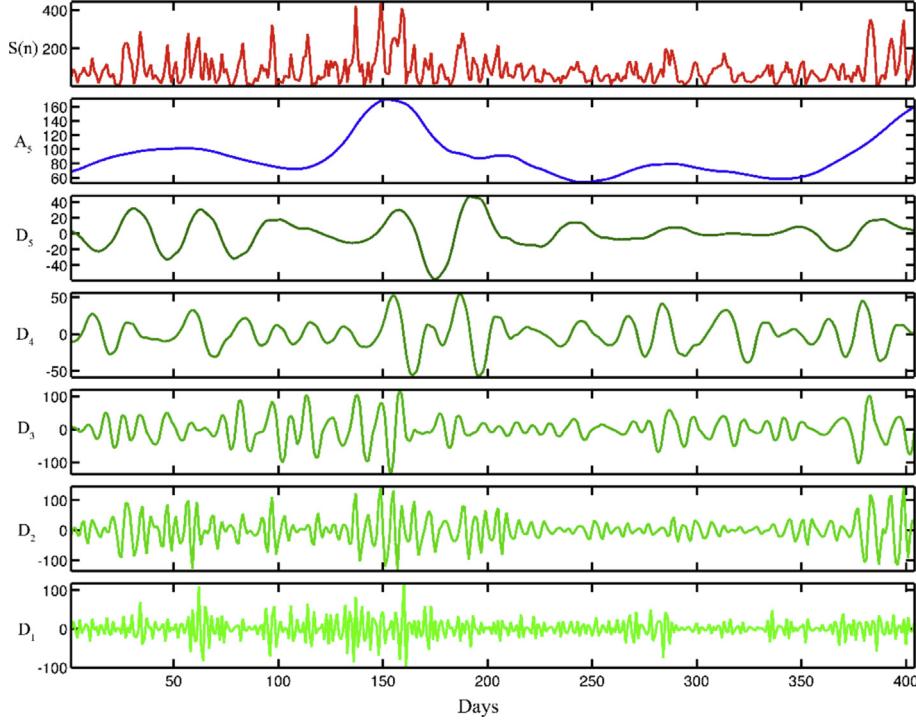


Fig. 4. The wavelet decomposition of the original time series $s(n)$ of PM_{2.5} concentrations at station D: D_1 – D_5 denote the wavelet coefficients at different levels and A_5 is the residual approximated signal of $s(n)$ on the fifth level.

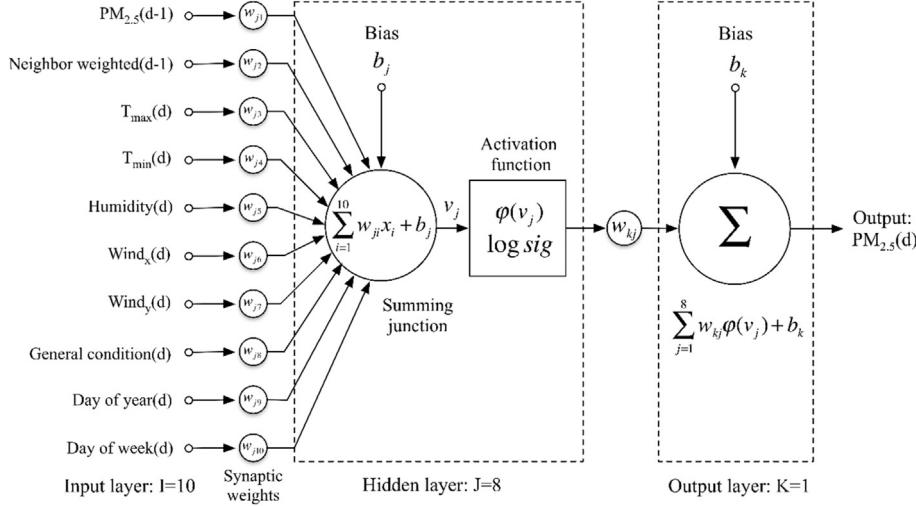


Fig. 5. The architecture of the MLP type of neural network (10-8-1) in this study.

one day in advance was then used as an input value for the next day's prediction together with other meteorological forecast data. More details about the ANN and MLP can be found in books (Bishop, 1995; Haykin, 1999).

All data were partitioned into two sets: 85% for training and 15% for testing. There are 10 variables in input layer, in which the prognostic predictors (temperature, humidity, wind) were used with their predicted values published by meteorological authorities instead of real-time value. All dataset were normalized in order to give similar impact of all input variables. After some trial and error, the neural network consisting of one hidden layer of 8 neurons (the structure 10-8-1) that performs best on the validation data (a 10% extraction from training data) was selected as the optimal MLP

structure. The model was built using the ANN toolbox in Matlab 2014a.

2.5. Error measure

To assess the performance of the models in a most objective way, two following methods have been used:

1. The error is usually stated as band error in some air quality websites. It represents the difference between the observed and forecasted intervals in which the observed and forecasted values fall. The range of pollution values is normally divided into five equal intervals. In our study, however, the maximum value of

$\text{PM}_{2.5}$ concentration reaches $448 \mu\text{g}/\text{m}^3$ due to the severe air pollution in Jing-Jin-Ji area, so a five intervals system cannot provide sufficient accuracy for the model assessment. Thus, we proposed a nine intervals system in this study: [0–50], [51–100], [101–150], ... [400–450]. This intervals system was chosen because it is the preferred method in air quality reporting (U.S. EPA, 2009). For example, the real and predicted pair (95, 105) for $\text{PM}_{2.5}$ is reported as +1 band error, since 95 falls into the second interval and 105 falls into the third interval. The same measure is used in Kurt and Oktay (2010), Domańska and Wojtyłak (2014).

2. Some widely used measures, including the mean absolute error (MAE), the root mean square error (RMSE), and the index of agreement (IA). They are defined by Eqs. (7)–(9):

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |O_i - P_i| \quad (7)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (O_i - P_i)^2} \quad (8)$$

$$\text{IA} = 1 - \frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (|O_i - \bar{O}| + |P_i - \bar{O}|)^2} \quad (9)$$

where N is the number of time points, O_i and P_i represent the observed and predicted values, and \bar{O} is the average of observed data.

3. Results and discussion

The aim of the experiments is to examine the effectiveness of the trajectory based geographic model and the wavelet

transformation used in ANN model for the prediction of $\text{PM}_{2.5}$ concentrations two days in advance. We have performed the experiments in three steps. In the first step, we used the ANN model alone (plain model) to make a direct approach. This plain model was treated as a reference model for the mixed approaches used in next two steps. Trajectory based geographic model and wavelet transformation were added to the plain model one by one in step 2 and 3. The architecture of ANN model described in Fig. 5 was used in all experiments except for step 1 (the neighbor weighted parameter was excluded). Particularly, the first input variable in ANN, $\text{PM}_{2.5}$ concentration, was substituted by the corresponding sub-series in step 3. An additional experiment has been conducted to test the capability of the models for predicting the $\text{PM}_{2.5}$ concentrations exceeding the air quality standard using the measurement of detection rate (DR) and false alarm rate (FAR).

At the beginning of experiments, we have divided the available data randomly into two sets: 344 points (85%) for training and 60 points (15%) for testing. To get the results in a most objective way, we have repeated 100 times the experiments of training and testing at randomly chosen composition of training and testing data. The final results of training and testing are the average of all trials. The results shown below are limited to the testing data except for the band error, because the percentage of $\pm n$ bands is not very stable on a limited testing set even though the cross validation has been used. Thanks to the early stopping method (Sarle, 1995) and “simpler-structure principle” (Jiang et al., 2004) used in ANN model, we avoid overfitting in training, which guarantees the similar results in training and testing.

The results of the experiments measured by band error for four selected stations in three steps are presented in Table 4. The best results are marked with bold. For plain model, the errors for +1 and +2 days are considered high for $\text{PM}_{2.5}$ forecasting, with the errors ~35% in ± 1 band, ~1.2% in ± 2 bands, and ~0.25% in ± 3 bands respectively. The best prediction is achieved for +1 day at station C. Better results are obtained by involving the trajectory based geographic model. The mixed approach yields lower error than the

Table 4
The results (band error) of different models for daily average forecasting values of $\text{PM}_{2.5}$.

Forecasting	Bands	Monitoring stations			
		A days (%)	B days (%)	C days (%)	D days (%)
Plain model					
+1 day	± 1 band	144 (35.64)	142 (35.15)	137 (33.91)	139 (34.41)
	± 2 bands	4 (1.00)	5 (1.24)	2 (0.50)	3 (0.74)
	± 3 bands	2 (0.50)	2 (0.50)	0 (0.00)	1 (0.25)
	Total	150 (37.14)	149 (36.89)	139 (34.41)	143 (35.40)
+2 day	± 1 band	149 (36.97)	145 (35.98)	140 (34.74)	141 (34.99)
	± 2 bands	5 (1.24)	7 (1.74)	4 (0.99)	6 (1.49)
	± 3 bands	2 (0.50)	1 (0.25)	0 (0.00)	0 (0.00)
	Total	156 (38.71)	153 (37.97)	144 (35.73)	147 (36.48)
Plain + Trajectory model					
+1 day	± 1 band	128 (31.68)	127 (31.44)	124 (30.69)	123 (30.45)
	± 2 bands	3 (0.74)	3 (0.74)	1 (0.25)	3 (0.74)
	± 3 bands	2 (0.50)	1 (0.25)	0 (0.00)	1 (0.25)
	Total	133 (32.92)	131 (32.43)	125 (30.94)	127 (31.44)
+2 day	± 1 band	135 (33.50)	133 (33.00)	128 (31.76)	130 (32.26)
	± 2 bands	4 (0.99)	3 (0.74)	2 (0.50)	3 (0.74)
	± 3 bands	2 (0.50)	2 (0.50)	0 (0.00)	0 (0.00)
	Total	141 (34.99)	138 (34.24)	130 (32.26)	133 (33.00)
Plain + Trajectory + Wavelet model					
+1 day	± 1 band	80 (19.80)	78 (19.31)	75 (18.56)	77 (19.06)
	± 2 bands	1 (0.25)	1 (0.25)	0 (0.00)	0 (0.00)
	± 3 bands	1 (0.25)	0 (0.00)	0 (0.00)	0 (0.00)
	Total	82 (20.30)	79 (19.56)	75 (18.56)	77 (19.06)
+2 day	± 1 band	84 (20.84)	83 (20.60)	76 (18.86)	72 (17.87)
	± 2 bands	2 (0.50)	1 (0.25)	0 (0.00)	2 (0.50)
	± 3 bands	1 (0.25)	1 (0.25)	0 (0.00)	0 (0.00)
	Total	87 (21.59)	85 (21.10)	76 (18.86)	74 (18.37)

Table 5

The results (RMSE, MAE, IA) of different models for daily average forecasting values of PM_{2.5}.

Monitoring station	Forecast measure	Plain model	Plain + trajectory model	Plain + trajectory + wavelet model
A	+1 day	RMSE	36.78	28.98
		MAE	27.84	21.52
		IA (%)	91.95	93.73
	+2 day	RMSE	38.59	30.10
		MAE	27.97	23.57
		IA (%)	89.92	92.40
B	+1 day	RMSE	34.54	28.35
		MAE	25.60	20.22
		IA (%)	92.38	94.08
	+2 day	RMSE	36.14	29.34
		MAE	27.72	22.76
		IA (%)	90.53	92.92
C	+1 day	RMSE	28.63	24.84
		MAE	20.94	19.21
		IA (%)	94.64	96.22
	+2 day	RMSE	30.72	26.41
		MAE	24.33	19.50
		IA (%)	93.01	95.29
D	+1 day	RMSE	31.71	26.62
		MAE	23.89	20.17
		IA (%)	94.67	96.29
	+2 day	RMSE	34.80	28.97
		MAE	25.87	21.58
		IA (%)	92.96	95.30

plain model, especially for ±1 (~32%) and ±2 (~0.7%) bands. The best prediction is also achieved for +1 day at station C. The results are quite satisfactory when the wavelet transformation is applied to the experiments. Errors have been significantly reduced on all bands, especially for ±2 (~0.2%), and ±3 (~0.1%) bands. The best prediction is achieved for +1 day at station C and +2 day at station D.

The results of the experiments measured by RMSE, MAE and IA for four selected stations in three steps are presented in Table 5. The best results are marked with bold. Similar with the results measured by band error, the forecast of the first day yields lower error than the second day. This can be explained by the theory of error accumulation, since the forecasting error for one day in advance is brought into the next day's prediction. Mixed approaches give better results than the direct approach (plain model) for both +1 and +2 days. It is seen that wavelet transformation plays an important role in obtaining good prediction results (~35% improvement on RMSE). From the aspect of absolute error (measured by RMSE and MAE), the best prediction is achieved for +1 day at station C using the hybrid model in step 3. The fact that station C yields lower absolute error than other stations in all trials indicates the importance of the local environment where a station locates. Station C locates in the zone of "clean source" (Fig. 3). The low variability of PM_{2.5} concentrations makes it easier to be predicted, as compared to other stations. Nevertheless, the best IA is achieved for both +1 and +2 days at station D, which

indicates that the relative measure can give an objective evaluation on the prediction model in different backgrounds. Due to the severe air pollution, the absolute errors are not so satisfactory as compared to other studies in developed counties, but the forecasting results show a good IA (~95%).

The capability of forecasting PM_{2.5} concentrations exceeding the air quality standard for the purpose of issuing warning, etc., is very crucial in air pollution forecasting system. According to the Ambient Air Quality Standards in China (EPA, 2012), it is considered that everyone may begin to experience adverse health effects when the PM_{2.5} concentration is greater than 115 µg/m³. Hence, we used the measurements of detection rate (the fraction of observed threshold exceedances predicted by the model) and false alarm rate (the fraction of predicted threshold exceedances in which the observed values are below the threshold) to check the usefulness of the models based on a 115 µg/m³ threshold in various tolerances. The results of the average values for four stations are shown in Table 6. The values of DR and FAR on +1 day forecast are better than those on +2 day at different tolerance levels. Hybrid models give better results than plain model, especially when wavelet transformation is applied. The increasing of tolerance makes the both indicators from good to excellent. The best result of DR (97.32%) and FAR (0.30%) is achieved for +1 day with the hybrid model in step 3 at the tolerance level of 20 µg/m³.

The scatter plots of observed and predicted daily average concentrations of PM_{2.5} during the whole experiment period at station

Table 6

Average detection rates (DR%) and false alarm rates (FAR%) for four monitoring stations with different models, based on a 115 µg/m³ threshold for various tolerances.

Forecast	Tolerance (µg/m ³)	Plain model		Plain + trajectory model		Plain + trajectory + wavelet model	
		DR (%)	FAR (%)	DR (%)	FAR (%)	DR (%)	FAR (%)
+1 day	0	71.43	3.77	76.47	4.45	85.71	2.74
	5	74.11	3.42	79.46	2.39	91.07	1.03
	10	76.79	3.08	84.82	2.05	93.75	0.89
	20	86.61	1.71	89.29	1.19	97.32	0.30
+2 day	0	66.07	4.81	70.76	3.44	81.25	3.28
	5	70.54	4.12	75.00	2.75	86.67	2.08
	10	75.89	3.09	83.93	2.13	91.18	1.72
	20	85.29	2.06	90.18	1.37	95.54	0.34

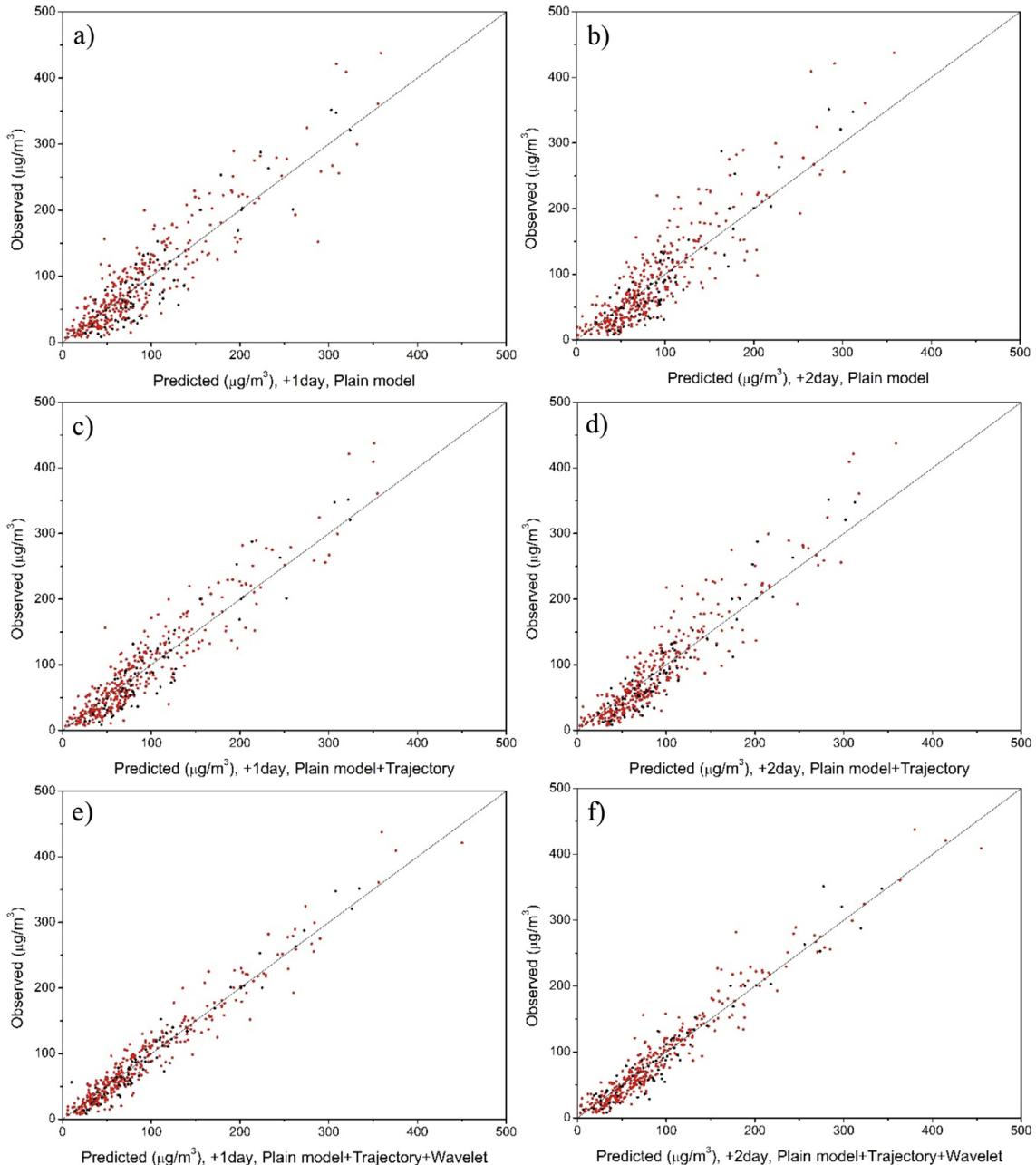


Fig. 6. Models performance using the training (red points) and testing (black points) dataset for the prediction of daily average concentrations of PM_{2.5} two days in advance at station D during the whole experiment period. The observed daily average concentrations of PM_{2.5} was compared with model performance from: a) +1 day plain model, b) +2 day plain model, c) +1 day plain model + trajectory, d) +2 day plain model + trajectory, e) +1 day plain model + trajectory + wavelet, f) +2 day plain model + trajectory + wavelet. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

D is illustrated in Fig. 6. The predicted values were calculated in one run of the experiment with the former 344 time points (one year's data) in training set (red points in Fig. 6) and the later 60 time points (two month data) in testing set (black points in Fig. 6). As we can see from all figures, points in both parts of training and testing show similar goodness-of-fit, which verifies the usefulness of the early stopping method and the “simpler-structure principle” on avoiding overfitting. Similar with the results concluded from the above tables, the predicted values of the first day follow closer to the changes of the pollution than the second day. Regarding the accuracy, mixed approaches are consistently far more accurate than plain model. Some of the high peaks missed by the plain model are almost anticipated by the hybrid model in step 3.

4. Conclusions

In this paper, a novel hybrid model is proposed that is capable to predict the daily average concentrations of PM_{2.5} two days in advance. It is built by applying the trajectory based geographic model and wavelet transformation into the MLP type of neural network. Combined with meteorological forecasts and respective pollutant predictors, the hybrid model is considered to be an effective tool to improve the forecasting accuracy of PM_{2.5}.

One significant novelty of this approach is the using trajectory based geographic parameter as an extra input predictor to the ANN model. This parameter calculated from the values of the selected three neighboring sites, is capable to capture both atmospheric and

geospatial information. The mixed approach outperforms the plain model and to some extent solves the problem of the under-prediction of high PM days.

Another novelty of this approach is the decomposition of the high variability time series into a few sub-series with lower variability and applying the prediction strategy to each of them at different scales. Application of this method has divided the prediction problem into a few simpler tasks allowing in this way to improve the forecasting accuracy.

The significant advantage of this hybrid model is the capability of predicting the high peaks of PM_{2.5} concentrations, which is considered a very critical factor in air pollution forecasting system. Due to the severe air pollution in Jing-Jin-Ji area, the overall accuracy of the prediction is not so satisfactory, as compared to other studies in developed counties. We believe the approach proposed here can be adapted to other regions and yields higher forecasting accuracy.

Acknowledgments

This work was supported by the National Key Technology R&D Program of the Ministry of Science and Technology of China (Grant No. 2012BAC20B06). We are very grateful for the two anonymous reviewers' insightful comments that significantly increase the clarity of this work.

References

- Alhanafy, T.E., Zaghlool, F., El, A.S., Moustafa, D., 2010. Neuro fuzzy modeling scheme for the prediction of air pollution. *J. Am. Sci.* 6, 605–616.
- Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Clarendon Press, Oxford.
- Brunelli, U., Piazza, V., Pignato, L., Sorbello, F., Vitabile, S., 2007. Two-days ahead prediction of daily maximum concentrations of SO₂, O₃, PM₁₀, NO₂, CO in the urban area of Palermo, Italy. *Atmos. Environ.* 41, 2967–2995.
- EPA, 2012. Technical Regulation on Ambient Air Quality Index (On Trial). Environmental Protection Administration of China. China Environmental Science Press.
- Chuang, M.-T., Zhang, Y., Kang, D., 2011. Application of WRF/Chem-MADRID for real-time air quality forecasting over the southeastern United States. *Atmos. Environ.* 45, 6241–6250.
- Daubechies, I., 1988. Ten Lectures on Wavelets. SIAM Press, Philadelphia, USA.
- Díaz-Robles, L.A., Ortega, J.C., Fu, J.S., Reed, G.D., Chow, J.C., Watson, J.G., Moncada-Herrera, J.A., 2008. A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: the case of Temuco, Chile. *Atmos. Environ.* 42, 8331–8340.
- Domańska, D., Wojtyłak, M., 2014. Explorative forecasting of air pollution. *Atmos. Environ.* 92, 19–30.
- Draxler, R.A., 1991. The accuracy of trajectories during ANATEX calculated using dynamic model analysis versus rawinsonde observations. *J. Appl. Meteorol.* 30, 1446–1467.
- Du, X., Kong, Q., Ge, W., Zhang, S., Fu, L., 2010. Characterization of personal exposure concentration of fine particles for adults and children exposed to high ambient concentrations in Beijing, China. *J. Environ. Sci.* 22, 1757–1764.
- Dutot, A.L., Rynkiewicz, J., Steiner, F.E., Rude, J., 2007. A 24-h forecast of ozone peaks and exceedance levels using neural classifiers and weather predictions. *Environ. Model. Softw.* 22, 1261–1269.
- Feng, X., Li, Q., Zhu, Y., Wang, J., Liang, H., Xu, R., 2014. Formation and dominant factors of haze pollution over Beijing and its peripheral areas in winter. *Atmos. Pollut. Res.* 5, 528–538.
- Gardner, M.W., Dorling, S.R., 1998. Artificial neural networks (the multilayer perceptron) – a review of applications in the atmospheric sciences. *Atmos. Environ.* 32, 2627–2636.
- Gardner, M.W., Dorling, S.R., 1999. Neural network modeling and prediction of hourly NO_x and NO₂ concentrations in urban air in London. *Atmos. Environ.* 33, 709–719.
- Genc, D.D., Yesilyurt, C., Tuncel, G., 2010. Air pollution forecasting in Ankara, Turkey using air pollution index and its relation to assimilate capacity of the atmosphere. *Environ. Monit. Assess.* 166, 11–27.
- Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene selection for cancer classification using support vector machines. *Mach. Learn.* 46, 389–422.
- Haykin, S., 1999. Neural Networks: a Comprehensive Foundation, second ed. Prentice Hall, Upper Saddle River, NJ, pp. 237–239.
- Hoi, K.I., Yuen, K.V., Mok, K.M., 2008. Kalman filter based prediction system for wintertime PM₁₀ concentrations in Macau. *Glob. NEST J.* 10, 140–150.
- Hooyberghs, J., Mensink, C., Dumont, G., Fierens, F., Brasseur, O., 2005. A neural network forecast for daily average PM₁₀ concentrations in Belgium. *Atmos. Environ.* 39, 3279–3289.
- Hrust, L., Klaić, Z.B., Križan, J., Antonić, O., Hercog, P., 2009. Neural network forecasting of air pollutants hourly concentrations using optimised temporal averages of meteorological variables and pollutant concentrations. *Atmos. Environ.* 43, 5588–5596.
- Jiang, D., Zhang, Y., Hu, X., Zeng, Y., Tan, J., Shao, D., 2004. Progress in developing an ANN model for air pollution index forecast. *Atmos. Environ.* 38, 7055–7064.
- Kukkonen, J., Partanen, L., Karppinen, A., Ruuskanen, J., Junninen, H., Kolehmainen, M., Niska, H., Dorling, S., Chatterton, T., Foxall, R., Cawley, G., 2003. Extensive evaluation of neural network models for the prediction of NO₂ and PM₁₀ concentrations, compared with a deterministic modeling system and measurements in central Helsinki. *Atmos. Environ.* 37, 4549–4550.
- Kuncheva, L., 2004. Combining Pattern Classifiers: Methods and Algorithms. Wiley, New York, USA.
- Kurt, A., Oktay, A.B., 2010. Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks. *Expert Syst. Appl.* 37, 7986–7992.
- Li, M., Hassan, R., 2010. Urban air pollution forecasting using artificial intelligence based tools. In: Villanyi, Vanda (Ed.), Air Pollution. In Tech, ISBN 978-953-307-143-5, pp. 195–219. Cht. 9.
- Liu, X., Li, J., Qu, Y., Han, T., Hou, L., Gu, J., Chen, C., Yang, Y., Liu, X., Yang, T., Zhang, Y., Tian, H., Hu, M., 2013. Formation and evolution mechanism of regional haze: a case study in the megacity Beijing, China. *Atmos. Chem. Phys.* 13, 4501–4514.
- Liu, Z., Liu, D., Huang, J., Vaughan, M., Uno, I., Sugimoto, N., Kittaka, C., Trepte, C., Wang, Z., Hostetler, C., Winker, D., 2008. Airborne dust distributions over the Tibetan Plateau and surrounding areas derived from the first year of CALIPSO lidar observations. *Atmos. Chem. Phys.* 8, 5045–5060.
- Lu, H.C., Hsieh, J.C., Chang, T.S., 2006. Prediction of daily maximum ozone concentrations from meteorological conditions using a two-stage neural network. *Atmos. Res.* 81, 124–139.
- Mallat, S., 1989. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. PAMI* 11, 674–693.
- Marshall, J.D., Nethery, E., Brauer, M., 2008. Within-urban variability in ambient air pollution: comparison of estimation methods. *Atmos. Environ.* 42, 1359–1369.
- McKeen, S., et al., 2009. An evaluation of real-time air quality forecasts and their urban emissions over eastern Texas during the summer of 2006 Second Texas Air Quality Study field study. *J. Geophys. Res. Atmos.* 114, D00F11.
- Niska, H., Rantamäki, M., Hiltunen, T., Karppinen, A., Kukkonen, J., Ruuskanen, J., Kolehmainen, M., 2005. Evaluation of an integrated modeling system containing a multi-layer perceptron model and the numerical weather prediction model HIRLAM for the forecasting of urban airborne pollutant concentrations. *Atmos. Environ.* 39, 6524–6536.
- Osowski, S., Garanty, K., 2007. Forecasting of the daily meteorological pollution using wavelets and support vector machine. *Eng. Appl. Artif. Intell.* 20, 745–755.
- Pérez, P., Reyes, J., 2002. Prediction of maximum of 24-h average of PM₁₀ concentrations 30 h in advance in Santiago, Chile. *Atmos. Environ.* 36, 4555–4561.
- Pérez, P., Reyes, J., 2006. An integrated neural network model for PM₁₀ forecasting. *Atmos. Environ.* 40, 2845–2851.
- Qiu, H., Yu, I., Wang, X., Tian, L., Tse, L.A., Wong, T.W., 2013. Differential effects of fine and coarse particles on daily emergency cardiovascular hospitalizations in Hong Kong. *Atmos. Environ.* 64, 296–302.
- Sarle, W.S., 1995. Stopped training and other remedies for overfitting. In: Proceedings of the 27th Symposium on the Interface of Computer Science and Statistics, pp. 352–360.
- Shad, R., Mesgari, M.S., Abkar, A., Shad, A., 2009. Predicting air pollution using fuzzy genetic linear membership Kriging in GIS. *Comput. Environ. Urban Syst.* 33, 472–481.
- Shaharuddin, M., Zaharim, A., Nor, M.J.M., Karim, O.A., Sopian, K., 2008. Application of wavelet transform on airborne suspended particulate matter and meteorological temporal variation. *WSEAS Trans. Top. Environ. Dev.* 4, 89–98.
- Shepherd, A.J., 1997. Second-order Methods for Neural Networks (New York).
- Siwek, K., Osowski, S., 2012. Improving the accuracy of prediction of PM₁₀ pollution by the wavelet transformation and an ensemble of neural predictors. *Eng. Appl. Artif. Intell.* 25, 1246–1258.
- Stadlober, E., Hörmann, S., Pfeifer, B., 2008. Quality and performance of a PM₁₀ daily forecasting model. *Atmos. Environ.* 42, 1098–1109.
- Stern, R., Buitjes, P., Schaap, M., Timmermans, R., Vautard, R., Hodzic, A., Memmesheimer, M., Feldmann, H., Renner, E., Wolke, R., Kerschbaumer, A., 2008. A model inter-comparison study focusing on episodes with elevated PM₁₀ concentrations. *Atmos. Environ.* 42, 4567–4588.
- Sun, W., Zhang, H., Palazoglu, A., Singh, A., Zhang, W., Liu, S., 2013. Prediction of 24-hour-average PM_{2.5} concentration using a hidden Markov model with different emission distributions in Northern California. *Sci. Total Environ.* 443, 93–103.
- U.S. EPA, 2009. Technical Assistance Document for Reporting of Daily Air Quality-air Quality Index. U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, North Carolina. EPA-454/B-09e001.
- Wang, J., Hu, M., Xu, C., Christakos, G., Zhao, Y., 2013. Estimation of citywide air pollution in Beijing. *PLOS One* 8, e53400.
- Zaharim, A., Shaharuddin, M., Nor, M.J.M., Karim, O.A., Sopian, K., 2009. Relationship between airborne particulate matter and meteorological variables using non-decimated wavelet transform. *Eur. J. Sci. Res.* 27, 308–312.
- Zhang, X., Wang, Y., Niu, T., Zhang, X., Gong, S., Zhang, Y., Sun, J., 2012. Atmospheric aerosol compositions in China: spatial/temporal variability, chemical signature,

- regional haze distribution and comparisons with global aerosols. *Atmos. Chem. Phys.* 12, 779–799.
- Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., Baklanov, A., 2012a. Real-time air quality forecasting, part I: history, techniques, and current status. *Atmos. Environ.* 60, 632–655.
- Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., Baklanov, A., 2012b. Real-time air quality forecasting, part II: state of the science, current research needs, and future prospects. *Atmos. Environ.* 60, 656–676.
- Zolghadri, A., Cazaurang, F., 2006. Adaptive nonlinear state-space modeling for the prediction of daily mean PM_{10} concentrations. *Environ. Model. Softw.* 21, 885–894.